

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:09:34 ; Search time 299.73 Seconds  
(without alignments)  
16.424 Million cell updates/sec

Title: US-09-331-631a-8\_COPY\_120\_161  
Perfect score: 245  
Sequence: 1 SORQPCQCHQHQEQRPE.....QQCVRECEKXYQENPWNGER 42

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP:REMBL\_15:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	111.5	45.5	525	10 Q43358	Q43358 theobroma c
2	103	42.0	593	10 Q9SEW4	Q9SEW4 juglans reg
3	98	40.4	625	10 Q9SP13	Q9SP13 macadamia i
4	98	40.0	666	10 Q9SP14	Q9SP14 macadamia i
5	94	38.4	666	10 Q9SP15	Q9SP15 macadamia i
6	91	37.1	393	10 Q9ZTP0	Q9ZTP0 oryza sativ
7	74	30.2	900	5 Q9VK09	Q9VK09 drosophila
8	74	30.2	3828	5 Q24742	Q24742 drosophila
9	72.5	29.6	402	10 Q9ZRH8	Q9ZRH8 oryza sativ
10	72	29.4	425	5 Q17400	Q17400 caenorhabd
11	72	29.4	600	5 Q17401	Q17401 caenorhabd
12	70	28.6	637	10 Q03678	Q03678 hordeum vul
13	70	28.6	905	5 Q18260	Q18260 caenorhabd
14	69.5	28.4	328	5 Q9VE66	Q9VE66 drosophila
15	69.5	28.4	925	5 Q9V7C5	Q9V7C5 drosophila
16	69	28.2	2123	5 Q9U9S7	Q9U9S7 dictyosteli
17	68	27.8	57	4 Q9Y4J1	Q9Y4J1 homo sapien
18	68	27.8	930	10 Q9S2Y9	Q9S2Y9 arabidopsis
19	67.5	27.6	709	5 Q9N4R2	Q9N4R2 caenorhabd

20	66.5	27.1	152	10 Q41168	Q41168 raphanus sa
21	66.5	27.1	174	10 Q41169	Q41169 raphanus sa
22	66	26.9	279	11 Q35846	Q35846 mus musculu
23	66	26.9	815	4 Q9UJG2	Q9UJG2 homo sapien
24	65.5	26.7	178	10 Q96339	Q96339 brassica ca
25	65.5	26.7	1457	5 Q9VT35	Q9VT35 drosophila
26	65	26.5	259	5 Q9M483	Q9M483 drosophila
27	65	26.5	388	5 Q16500	Q16500 caenorhabd
28	65	26.5	388	5 Q16501	Q16501 caenorhabd
29	65	26.5	419	5 Q9NA38	Q9NA38 caenorhabd
30	65	26.5	420	5 Q9NA61	Q9NA61 caenorhabd
31	65	26.5	438	5 Q16502	Q16502 caenorhabd
32	65	26.5	445	5 Q16511	Q16511 caenorhabd
33	64.5	26.3	411	5 P91419	P91419 caenorhabd
34	64.5	26.3	425	5 Q15755	Q15755 dictyosteli
35	64	26.1	129	2 Q9RNE4	Q9RNE4 streptococc
36	64	26.1	242	5 Q19919	Q19919 caenorhabd
37	64	26.1	549	5 Q9VF63	Q9VF63 drosophila
38	64	26.1	687	3 P87065	P87065 candida alb
39	64	26.1	838	5 Q27422	Q27422 caenorhabd
40	64	26.1	838	5 Q18761	Q18761 caenorhabd
41	64	26.1	1068	5 Q9W539	Q9W539 drosophila
42	64	26.1	1464	5 Q96836	Q96836 drosophila
43	64	26.1	3583	5 Q9V3N4	Q9V3N4 drosophila
44	64	26.1	3584	5 Q9U999	Q9U999 drosophila
45	63.5	25.9	101	5 Q9TVK8	Q9TVK8 drosophila

## ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	525 AA.
Q43358	Q43358			
AC	Q43358;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	VICILIN PRECURSOR.			
GN	CSV.			
OS	Theobroma cacao (Cacao).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;			
OC	Malvales; Malvaceae; Theobroma.			
OX	NCBI_TaxID=3641;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LEAVES;			
RX	MEDLINE=92288309; PubMed=1600151;			
RA	McHenry L., Fritz P.J.;			
RT	*Comparison of the structure and nucleotide sequences of vicilin genes			
RT	of cocoa and cotton raise questions about vicilin evolution.*;			
RL	Plant Mol. Biol. 18:1173-1176(1992).			
DR	EMBL: X62625; CAA4493.1; -			
DR	EMBL: X62626; CAA4494.1; -			
DR	HSSP: P02853; 2PHL.			
DR	MENDEL: 30919; Thecc:1188;30919.			
DR	INTERPRO: IPR001113; -			
DR	PFAM: PF00546; Seedstore-7s; 1.			
DR	PRODOM: PD081059; -; 1.			
KW	SIGNAL.			
FT	SIGNAL			
FT	CHAIN			
SO	SEQUENCE			

Query Match 45.5%; Score 111.5; DB 10; Length 525;  
Best Local Similarity 50.0%; Pred. No. 6.8e-06;  
Matches 20; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 2 SORQPCQCHQHQEQRPEKKQCVRECEKXYQENPWNGER 41  
111:111 1:111:111 1:111 1:111 111  
DB 82 SORQPCQCHQHQEQRPEKKQCVRECEKXYQENPWNGER 120

```

RESULT 2
Q9SEW4 PRELIMINARY: PRT: 593 AA.
ID AC Q9SEW4:
AC Q9SEW4:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE VICILIN-LIKE PROTEIN PRECURSOR (FRAGMENT).
OS Juglans regia (English walnut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fagales; Juglandaceae; Juglans.
OX NCBI_TaxID=51240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SUNLAND; TISSUE=SONATIC EMBRYO LINE;
RA Teuber S.S., Jarvis K.C., Peterson W.R., Dandekar A.M., Ansari A.A.;
RT "Identification and cloning of a cDNA encoding a vicilin-like protein,
RT allergen.",
RT J. Biol. Chem. 271:11111-11118 (1996).
RT PubMed: 8711111.
RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF066055; AAF18269.1; -.
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; -.
DR PFM: PF00546; Seedstore_7s: 1.
DR NON_TER 1
SQ SEQUENCE 593 AA: 69990 MW: 98A127E19B18C0D8 CRC64;

Query Match 42.0%; Score 103; DB 10; Length 593;
Best Local Similarity 42.1%; Pred. No. 8e-05;
Matches 16; Conservative 9; Mismatches 13; Indels 0; Gaps 0;

QY 2 OROFECQCHQOORPEKKOOCVRECKEYQOE-----NPMRG 39
DB 120 QOYHRCORRCOIQOESPOROKOOCORCERYKQOGR 157

RESULT 3
Q9SPL3 PRELIMINARY: PRT: 625 AA.
ID AC Q9SPL3:
AC Q9SPL3:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE VICILIN PRECURSOR (FRAGMENT).
CN AMP2.
OS Macadamia integrifolia (Macadamia nut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.
OX NCBI_TaxID=60698;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NUT KERNEL;
RA Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;
RT "A family of antimicrobial peptides is produced by processing of a 7S
RT globulin protein in Macadamia integrifolia.",
RT Plant J. 0:0-0(1999).
RL EMBL: AF161885; AAD54246.1; -.
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; -.
DR PFM: PF00546; Seedstore_7s: 1.
DR NON_TER 1
SQ SEQUENCE 625 AA: 73586 MW: 415808A89D370296 CRC64;

Query Match 40.4%; Score 99; DB 10; Length 625;
Best Local Similarity 38.3%; Pred. No. 0.00025;
Matches 18; Conservative 11; Mismatches 8; Indels 10; Gaps 2;

QY 2 OROFECQCHQOORPEKKOOCVRECKEYQOE-----NPMRG 40

```

```

DB 150 QREYEDCRRRCEQOE--PRLOYOCORRCOORORHGRGCDLNNPORG 194

RESULT 4
Q9SPL4 PRELIMINARY: PRT: 666 AA.
ID AC Q9SPL4:
AC Q9SPL4:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE VICILIN PRECURSOR.
CN AMP2.
OS Macadamia integrifolia (Macadamia nut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.
OX NCBI_TaxID=60698;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NUT KERNEL;
RA Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;
RT "A family of antimicrobial peptides is produced by processing of a 7S
RT globulin protein in Macadamia integrifolia.",
RT Plant J. 0:0-0(1999).
RL EMBL: AF161884; AAD54245.1; -.
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; -.
DR PFM: PF00546; Seedstore_7s: 1.
DR NON_TER 1
SQ SEQUENCE 666 AA: 78243 MW: 0ECA22F8710F8A7B CRC64;

Query Match 40.0%; Score 98; DB 10; Length 666;
Best Local Similarity 38.3%; Pred. No. 0.00035;
Matches 18; Conservative 11; Mismatches 8; Indels 10; Gaps 2;

QY 2 OROFECQCHQOORPEKKOOCVRECKEYQOE-----NPMRG 40
DB 191 QREYEDCRRRCEQOE--PROYOCORRCOORORHGRGCDLNNPORG 235

```

```

RESULT 5
Q9SPL5 PRELIMINARY: PRT: 666 AA.
ID AC Q9SPL5:
AC Q9SPL5:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE VICILIN PRECURSOR.
CN AMP2.
OS Macadamia integrifolia (Macadamia nut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.
OX NCBI_TaxID=60698;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NUT KERNEL;
RA Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;
RT "A family of antimicrobial peptides is produced by processing of a 7S
RT globulin protein in Macadamia integrifolia kernels.",
RT Plant J. 0:0-0(1999).
RL EMBL: AF161883; AAD54244.1; -.
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; -.
DR PFM: PF00546; Seedstore_7s: 1.
DR NON_TER 1
SQ SEQUENCE 666 AA: 78217 MW: C752B884B2DF0224 CRC64;

Query Match 38.4%; Score 94; DB 10; Length 666;
Best Local Similarity 35.3%; Pred. No. 0.0011;
Matches 12; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 2 OROFECQCHQOORPEKKOOCVRECKEYQOE 35

```

Page 3

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Moharry C., Morris J., Moshfeghi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclel J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpkins M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svaitkas R., Tector C., Turner R., Venier E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,  
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao O.-A.,  
RA Ye J., Yeh R.-F., Zaveri J., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,"  
RT "The genome sequence of *Drosophila melanogaster*.";  
RT Science 287:2185-2195(2000).  
RL EMBL: AEO03639; AAF5374.1; -  
DR FLYBASE: FBgn0032497; CG6043.  
DR INTERPRO: IPR002965; -  
DR PRINTS: PR01217; PRICHEXENSN.  
SQ SEQUENCE 900 AA: 89948 MW: DD99C0057CDA5E44 CRC64;

Best Local Similarity 39.0%; Pred. No. 0.36; Length 200,  
Matches 16; Conservative 10; Mismatches 13; Indels 2; Gaps 1

[illegible]



RESULT	14		
Q9VE66			
1D	Q9VE66	PRELIMINARY;	PRT; 328 AA.
AC	Q9VE66;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)		
DE	CG14307 PROTEIN.		
GN	CG14307.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		

CC Phlebotominae: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:  
OC Ephydroidea: Drosophilidae: Drosophila.  
OX NCBI\_taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.C., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Branton R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Celter A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Dudkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwan C.,  
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kerchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svayrtkas R., Teclor C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhu Q., Zheng L.,  
RA Zheng X.H., Zhong F., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA GIBBS R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003722; AAF55563.1; -  
DR HSSP: P17789; ZDRP.  
DR FLYBASE: FBgn0038624; CG14307.  
DR INTERPRO: IPR000822; -  
DR PFM: PF00006; zfc2H2. 1.  
DR PROSITE: PS00026; ZINC\_FINGER\_C2H2. 1.  
OX SEQUENCE 328 AA; 38204 MW; 4C169889ABFD0899 CRC64;

[illegible]

